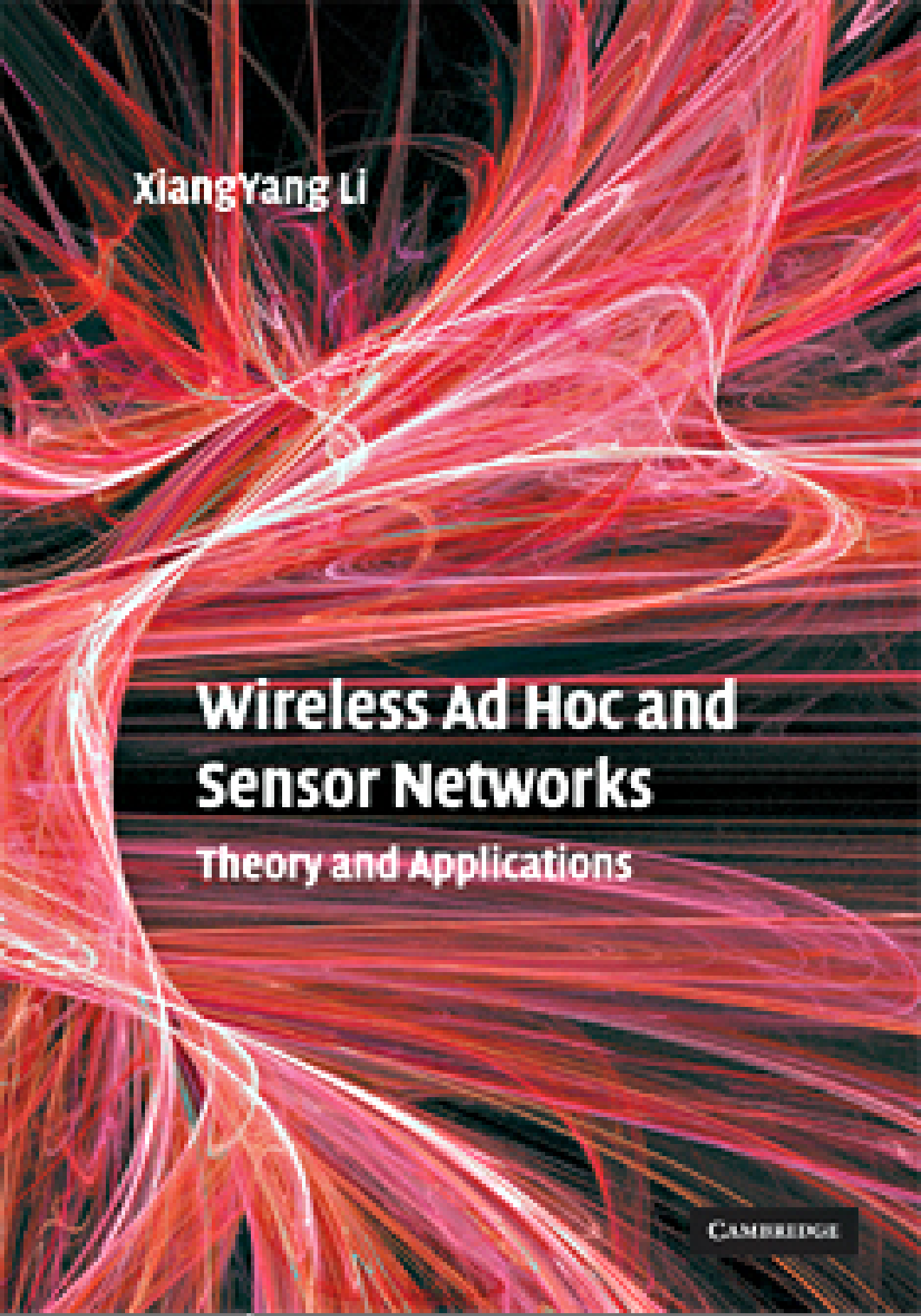


CAMBRIDGE www.cambridge.org/9780521865234

This page intentionally left blank

Wireless Ad Hoc and Sensor Networks

Wireless Ad Hoc and Sensor Networks describes the theory of ad hoc networks. It also demonstrates techniques for designing efficient algorithms and systematically analyzing their performance.

Li develops the fundamental understanding required to tackle problems in these networks by first reviewing relevant protocols, then formulating problems mathematically, and solving them algorithmically. Wireless MAC protocols, including various IEEE 802.11 protocols, 802.16, Bluetooth, and protocols for wireless sensor networks are treated in detail. Channel assignment for maximizing network capacity is covered; topology control methods are explored at length; and routing protocols for unicast, broadcast, and multicast are described and evaluated. Cross-layer optimization is also considered.

The result is a detailed account of the various algorithmic, graph-theoretical, computational-geometric, and probabilistic approaches to attack problems faced in these networks, delivering an understanding that will allow readers to develop practical solutions for themselves. This title is an invaluable resource for graduate students and researchers in electrical engineering and computer science departments, as well as for practitioners in the communications industry.

XiangYang Li is currently an associate professor of computer science at the Illinois Institute of Technology. He also holds a visiting professorship or adjunct-professorship at TianJing University, WuHan University, and NanJing University, in China. He was awarded his Ph.D. in 2001 from the Department of Computer Science at the University of Illinois at Urbana-Champaign. A leading researcher in the field of wireless networks, he has made important contributions in the areas of network topology and routing. His current research interests include cooperation, energy efficiency, and distributed algorithms for wireless ad hoc and sensor networks.

Wireless Ad Hoc and Sensor Networks

Theory and Applications

XIANGYANG LI
Illinois Institute of Technology

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521865234

First published in print format 2008

ISBN-13 978-0-511-42139-6 eBook (Adobe Reader)

ISBN-13 978-0-521-86523-4 hardback

To my wife, Min
my daughter, Sophia
my son, Kevin
and my families

Contents

Preface

Introduction

In the next generation of wireless communication systems, there will be a need for the rapid deployment of independent mobile users. Significant examples include establishing survivable, efficient, dynamic communication for emergency/rescue operations, disaster relief efforts, and military networks. Such network scenarios cannot rely on centralized and organized connectivity and can be conceived as applications of mobile ad hoc networks (MANETs). A MANET is an autonomous collection of mobile users that communicate over relatively bandwidth-constrained wireless links. Because the nodes are mobile, the network topology may change rapidly and unpredictably over time. The network is decentralized; all network activity, including discovering the topology and delivering messages, must be executed by the nodes themselves; that is, routing functionality will be incorporated into mobile nodes.

In many commercial and industrial applications, we often need to monitor the environment and collect the information about the environment. In some of these applications, it would be difficult or expensive to monitor using wired sensors. If this is the case, wireless sensor networks in which sensors are connected by wireless networks are preferred. A wireless sensor network (WSN) consists of a number of sensors spread across a geographic area. Each sensor node has wireless communication capability and some level of intelligence for signal-processing and networking of data. A WSN could be deployed in wilderness areas for a sufficiently long time (e.g., years) without the need to recharge or replace the power supplies. Typical applications of WSNs include monitoring, tracking, and controlling.

The subject of wireless ad hoc networking and sensor networking is enormously complex, involving many concepts, protocols, technologies, algorithms, and products that work together in an intricate manner. The set of applications for MANETs is diverse, ranging from small, static networks that are constrained by power sources to large-scale, mobile, highly dynamic networks. Recently, wireless sensor networks have also been used in Supervisory Control and Data Acquisition (SCADA). SCADA systems are used to monitor or to control chemical or transport processes, in municipal water supply systems, control electric power generation, transmission, and distribution, gas and oil pipelines, and other distributed processes. The design of network protocols for these networks is a complex issue. Regardless of the application, MANETs and sensor networks need efficient distributed algorithms determining network organization, linking

scheduling, and routing. However, determining feasible routing paths and delivering messages in a decentralized environment in which network topology fluctuates is not a well-defined problem. Although the shortest path (based on a given cost function) from a source to a destination in a static network is usually the optimal route, this idea is not easily extended to MANETs. Factors such as variable wireless link quality, propagation path loss, fading, multiuser interference, power expended, and topological changes become relevant issues. The network should be able to adaptively alter the routing paths to alleviate any of these effects. Moreover, in a military environment, preservation of security, latency, reliability, intentional jamming, and recovery from failure are significant concerns. Military networks are designed to maintain a low probability of intercept and/or a low probability of detection. Hence, nodes prefer to radiate as little power as necessary and transmit as infrequently as possible, thus decreasing the probability of detection or interception. A lapse in any of these requirements may degrade the performance and dependability of the network.

The basic goals of a wireless ad hoc sensor network generally depend on the application, but the following tasks are common to many networks:

1. *Determine the value of some parameter at a given location*: In an environmental network, one might want to know the temperature, atmospheric pressure, amount of sunlight, and relative humidity at a number of locations. This example shows that a given sensor node may be connected to different types of sensors, each with a different sampling rate and range of allowed values.
2. *Detect the occurrence of events of interest and estimate parameters of the detected event or events*: In the traffic sensor network, one would like to detect a vehicle moving through an intersection and estimate the speed and direction of the vehicle.
3. *Classify a detected object*: Is a vehicle in a traffic sensor network a car, a minivan, a light truck, a bus, and so on?
4. *Track an object*: In a military sensor network, one would like to track an enemy tank as it moves through the geographic area covered by the network.

In these four tasks, an important requirement of the sensor network is that the required data be disseminated to the proper end users. In some cases, there are fairly strict time requirements on this communication. For example, the detection of an intruder in a surveillance network should be immediately communicated to the police so that action can be taken. Because wireless sensors are often powered by batteries only, energy efficiency is critical for the lifetime of a wireless sensor network. Thus, a considerable amount of research has recently been devoted to developing energy-efficient protocols for wireless sensor networks. In addition to energy-efficient protocols, wireless ad hoc sensor network requirements include but are not limited to scalability (to support a large number of mostly stationary sensors for which networks of 10,000 or even 100,000 nodes are envisioned), network self-organization to support scalability and fault tolerance, collaborative signal-processing, and querying ability. Given the large number of nodes and their potential placement in hostile locations, it is essential that the network be able to self-organize; manual configuration is not feasible. Moreover, nodes may fail

(either from lack of energy or from physical destruction), and new nodes may join the network. Therefore, the network must be able to periodically reconfigure itself so that it can continue to function. Individual nodes may become disconnected from the rest of the network, but a high degree of connectivity must be maintained. Another factor that distinguishes wireless sensor networks from MANETs is that the end goal is detection/estimation of some events of interest and not just communications. To improve the detection/estimation performance, it is often quite useful to fuse data from multiple sensors. This data fusion requires the transmission of data and control messages, and so it may put constraints on the network architecture. A user may want to query an individual node or a group of nodes for information collected in the region. Depending on the amount of data fusion performed, it may not be feasible to transmit a large amount of the data across the network. Instead, various local sink nodes will collect the data from a given area and create summary messages. A query may be directed to the sink node nearest the desired location.

Recent years have seen a great amount of research in wireless networks, especially wireless ad hoc networks. These works involve a number of theoretical aspects of computer science, including approximation algorithms, computational geometry, combinatorics, and distributed algorithms. Because of the limited capability of processing power, storage, and energy supply, many conventional algorithms are too complicated to be implemented in wireless ad hoc and sensor networks. Some other algorithms do not take advantage of the geometric nature of the wireless networks. Additionally, most of the currently developed location-based algorithms for wireless networks assume a precise position of each wireless node, which is impossible practically. The majority of the algorithms with theoretical performance guarantee developed in this area also assume that all nodes have a uniform transmission range. These algorithms will likely fail when nodes have disparate transmission ranges. In summary, the wireless ad hoc and sensor networks require efficient distributed algorithms with low computation complexity, low communication complexity, and low storage complexity. These algorithms are expected to take advantage of the geometry nature of the wireless ad hoc networks. Several fundamental questions should be answered: Can we improve the performance of traditional distributed algorithms, developed for wired networks, under wireless ad hoc networks? Does the position information of wireless nodes make a difference in algorithm performance? Much of the existing work in wireless ad hoc networking also assumes that each individual wireless node (possibly owned by selfish users) will follow prescribed protocols without deviation. However, each user may modify the behavior of an algorithm for self-interest reasons. How are desired global-system performances achieved when individual nodes are selfish?

This is a new book aimed at the teaching of wireless ad hoc and sensor networks from the algorithmic and theoretical perspective. The primary focus of the book is on the algorithms, especially efficient distributed algorithms, related wireless ad hoc protocols, and some fundamental theoretical studies of phenomena in wireless ad hoc and sensor networks. Many aspects of wireless networking are covered at the introductory level. I tried to cover as many interesting and algorithmic challenging topics related to wireless ad hoc and/or sensor networks as possible in this book. I know that several interesting

topics and elegant algorithms are missing. Some are due to lack of space and some are due to the theme of the book. No judgment is implied for algorithms and protocols not covered in this book.

Audience

This book is intended for graduate students, researchers, and practitioners who are interested in obtaining a detailed overview of a number of various algorithmic, graph-theoretical, computational-geometric, and probabilistic approaches to attack certain challenging problems stemming from wireless networks, especially wireless ad hoc and sensor networks. Thus, when I wrote this book, I tried to cover many details for most of the algorithms studied. This book can, in general, serve as a reference resouce for researchers, engineers, and protocol developers working in the field of wireless ad hoc and/or sensor networks. Consequently, most of the chapters are written in such a way that they can be read and taught independently.

While I have tried to make the book (and most chapters) as self-contained as possible, some rudimentary knowledge of algorithm design and analysis, computational geometry, distributed systems, graph theory, linear algebra, networking protocols, and probability theory is required for reading this book.

Organization of the Book

This book essentially is organized based on the layers of wireless networking: the physical and medium-access-control (MAC) layers, the topology control functions that lie between the MAC and network routing layer, and the network routing layer.

The first part of the book presents introductory material that is necessary for the rest of the book.

Chapter 1 briefly reviews the history of wireless communications and discusses different wireless networks, such as infrastructure-based wireless networks (cellular networks) and infrastructureless wireless networks. Among infrastructureless networks, wireless ad hoc networks and wireless sensor networks are briefly discussed.

Chapter 2 covers some fundamentals of wireless transmissions. In this chapter, we study the interference constraints of wireless communications, the wireless propagation model, and the channel capacity of a wireless channel. We also define the communication graph and the interference graph (or conflict graph) induced by a wireless network. Because minimizing energy consumption is critical for the success of many wireless networks, we also review several energy-consumption models that are often used in the literature. Additionally, we discuss a number of mobility models to simulate mobile networks.

The second part of the book is mainly about the MAC protocols for wireless networks. We study CSMA, TDMA, and CDMA protocols.

Chapter 3 concentrates on the CSMA-based wireless MAC protocols. We study how hidden-terminal and exposed-terminal problems are addressed. We also briefly study several typical wireless MAC protocols such as IEEE 802.11 (or WiFi) protocols for wireless LANs, IEEE 802.16 (WiMAX) for mesh networks, and Bluetooth for wireless personal area networks. We briefly review some of the specific MAC protocols proposed for wireless sensor networks that integrate CSMA and TDMA.

Chapter 4 concentrates on the MAC protocols based on TDMA. These protocols assume that the time is slotted and that each link will be assigned some time slots, in which it can transmit data over this link. When a link is assigned a time slot, it is guaranteed that no wireless interference will occur when it uses this link at this time slot. This assignment is often 0/1: A slot either is assigned to a link or is not assigned. When a time slot is not assigned, a link cannot transmit at that specific time slot. We study some TDMA-based link-scheduling algorithms that can provide theoretical performance guarantees.

Chapter 5 concentrates on spectrum channel assignment for wireless networks (cellular networks and wireless ad hoc networks). We first study how to assign channels for a set of access networks such that the network capacity is maximized, or the number of assigned channels is minimized while certain capacity requirements are satisfied. We then study the results for spectrum channel assignment for ad hoc networks. The objective of a channel assignment could be to use the least number of channels to achieve a connected network while the channel availability and network interface constraints at all nodes are satisfied. We also study the transition phenomena of a number of network properties depending on the availability of a wireless spectrum.

Chapter 6 studies several algorithms for assigning a CDMA code to wireless networks when CDMA is supported.

The third part of the book is about topology control and power assignment for wireless networks.

Chapter 7 studies the construction of backbone for wireless networks. Backbone is especially useful for routing in mobile networks. We study several centralized and distributed algorithms that can construct a network backbone (i.e., a connected dominating set) whose size is within a constant factor of the optimum for wireless networks modeled by a unit disk graph. We also study some pure localized algorithms that have lower communication costs, although the theoretical constant-approximation ratio on the backbone size is not guaranteed.

Chapter 8 studies the construction of a backbone network when each wireless node has a weight denoting its cost of being at the backbone. The objective is to minimize the total weight of the backbone. We study several algorithms with good approximation ratios.

Chapter 9 studies topology-control algorithms that will construct flat network topologies with proved performance guarantees. Here, a network topology is said to be flat if every node in the network will assume the same role in network routing. Notice that for a backbone-based structure, the node on the backbone will forward the messages for nodes that are not on the backbone. We study efficient distributed algorithms that can construct energy-efficient network topologies.

Chapter 10 studies the power-assignment problems for wireless networks. Power assignment is selecting a transmission power for each node in the network such that the resulting communication network using the allocated transmission power has certain properties. The objective of a power assignment is often to minimize the total power used by all nodes or to minimize the maximum transmission power of all nodes. The latter is often easy to solve, based on a binary search on all choices of transmission power. We study algorithms that assign transmission powers such that the network is connected, k-connected, or consumes the least power for broadcast or multicast.

Chapters 11 and 12 are related to previous chapters but with different focuses. In these two chapters, we study the so-called transition phenomena of random wireless networks; in other words, the behavior of some certain parameters of the network when the number of nodes in the network goes to infinity. In Chapter 11, we study the critical transmission range r_n when a random network of n nodes distributed in a given region (typically, a unit square or a disk with unit area) is connected with high probability or k-connected with high probability. In Chapter 12, we study the critical node degree needed for producing a connected random network with high probability; the critical transmission range for connectivity in sparse networks or in mobile networks; the critical transmission range for a successful routing with high probability for certain localized routing algorithms; and the critical sensing range for covering a region with high probability.

The fourth part of the book is on routing protocols for wireless networks. We study routing protocols for unicast, multicast, and broadcast, and routing protocols with selfish agents.

Chapter 13 studies the energy-efficient unicast routing for wireless networks. We briefly review some typical proactive and reactive unicast routing protocols proposed in the literature, such as DSDV, OLSR, AODV, DSR, and opportunistic routing. We also study geographic routing protocols that utilize the geometry information of wireless nodes to improve the routing performance. Cluster-based hierarchical routing is also briefly discussed.

Chapter 14 studies energy-efficient routing protocols for broadcast and multicast. We first study some centralized algorithms for energy-efficient broadcast and multicast. These algorithms are often based on the node-weighted or link-weighted Steiner tree algorithms proposed in the literature. Later, we study several distributed or localized methods that are practically efficient.

Chapter 15 studies the routing from another point of view. In all previous protocols, it has been assumed that all wireless nodes will follow predescribed protocols without deviation. In practice, this may not be true, especially when wireless nodes are owned by individuals. In this chapter, we study how to design routing protocols when we know that individual nodes may not follow a routing protocol for their own benefit. We study this problem mainly using a game-theoretical approach, although a number of different approaches are also briefly discussed. In the game-theoretical-based approaches, wireless nodes will be compensated for their services to others. We study how each individual relay node is paid and how the payment to these nodes is implemented. For multicast, we also study algorithms that will fairly share the payments to relay nodes among potential receivers.

Chapter 16 studies how to improve the network through a cross-layer approach of jointly optimizing routing, link scheduling, and channel assignment. We formulate this problem as mixed-integer programming and then relax it to linear programming. By combining it with link scheduling, we show that the relaxed linear-programming formulation will find a solution that is at least a constant factor of the optimum for a number of network models.

The fifth and the last part of the book is devoted to studying a few other interesting topics in wireless networks; for example, location tracking, the performance of random networks, and security.

Chapter 17 studies finding the location of wireless sensor nodes and tracking the position of a moving object by using wireless sensor networks.

In previous chapters, especially Chapter 16, we study what maximum throughput is achievable by a given wireless network under a certain wireless interference model.

Chapter 18 concentrates on the asymptotic network capacity of a random network. We study how the capacity of wireless networks scale with the number of nodes in the networks (when given a fixed deployment region) or scale with the size of the deployment region (when given a fixed deployment density) for a number of operations, such as unicast and broadcast. We especially study a pioneering work by Gupta and Kumar on the network capacity of a random network for unicast. We also study the network capacity for broadcast under various channel models.

Chapter 19 concentrates on ensuring security in wireless networks. We mainly focus on some fundamentals of cryptography, some key-predistribution protocols, and some secure routing protocols proposed in the literature. Cryptography will provide us some fundamental tools such as symmetric-key and asymmetric-key encryption, digital signature, and hash functions to implement some security protocols. We then review some secure routing protocols proposed in the literature.

Acknowledgments

This work is supported in part by the National Basic Research Program of China (973 Program) under Grant No. 2006CB303000 and the National High Technology Research and Development Program of China (863 Program) under Grant No. 2007AA01Z180.

First, this book is in memory of Professor Chao-Ju (Jennifer) Hou from UIUC, a great researcher, an active leader, and a close colleague, who passed away due to breast cancer on December 2, 2007. I am indebted to her and many colleagues for background material and insightful input, which, in one way or another, have helped me to write certain parts of the book. I am also deeply grateful to many colleagues and peer researchers who shared with me in recent years the exciting tasks of studying algorithms and protocols with theoretical performance guarantees for wireless networks, whose research results have helped me in writing several chapters, and who provided feedback to early versions of this book. Especially, I would like to thank the following: Stefano Basagni, Gruia Călinescu, Guohong Cao, Marco Conti, Stephan Eidenbenz, Ophir Frieder, Jie Gao, Jennifer Hou, Jean-Pierre Hubaux, XiaoHua Jia, Sanjiv Kapoor, P. R. Kumar, Baochun Li, Erran Li, Qun Li, Xin Liu, YunHao Liu, Songwu Lu, Thyaga Nandagopal, Paolo Santi, Ivan Stojmenovic, Nitin Vaidya, Peng-Jun Wan, Roger Wattenhofer, Jie Wu, GuoLiang Xue, Frances Yao, and Chih-Wei Yi. I also would like to thank my (former and current) Ph.D. students at Illinois Institute of Technology: Professor Yu Wang, Dr. Weizhao Wang, Professor WenZhan Song, Kousha Moaveninejad, YanWei Wu, XuFei Mao, Ping Xu, Shajojie Tang, and XiaoHua Xu. Much of the material presented in this book is the fruit of our collaboration during the past few years.

The continued support and encouragement of my dad, KaiWen Li; my mom, LanHua Wu; my wife, Min Chen; my lovely daughter, Sophia Li; and my bright son, Kevin Li were essential ingredients in the completion of the book. I am grateful to them all. Without their unwearying love and patience, tireless encouragement, and full unconditional support, this book would never have been possible. I am also grateful for all my friends and colleagues who provided extraordinary insights, gave me helpful advice in research and life, and pointed me in the right direction.

Last but by no means least, I am deeply grateful to Anna Littlewood, Cambridge University Press, for her tireless effort in all matters relating to the preparation and production of many different versions of the manuscript for the book. I am also deeply grateful to Victoria Danahy and Barbara Walthall of Aptara Corp. for carefully copy-editing all chapters of the book.

I am sure that I have missed many others, although not intentionally; I thank all of you.

XiangYang Li
Chicago, Illinois
February 2008

Abbreviations

1D	one-dimensional
2D	two-dimensional
2G	second-generation
3D	three-dimensional
3G	third-generation
ABR	associativity-based routing
ACK	acknowledgment (frame)
A/D	analog-to-digital (conversion)
AES	Advanced Encryption System
AFR	adaptive face routing
AHLoS	ad hoc localization system
AIFS	arbitration interframe space
Algorithm KV	algorithm of Khuller and Viskhin
AMPS	Advanced Mobile Telephone System
amp	amplifier
AoA	angle of arrival
AODV	ad hoc on-demand vector (routing)
AP	access point
APS	ad hoc positioning system
APX	approximable
APXH	APX-hard
AS	autonomous system
ATIM	ad hoc traffic indication map
ATM	asynchronous transfer mode
AWA	Accessos Web Alternativos
BAIP	broadcast average incremental power
BB	budget balance
BFS	breadth first search
BGP	border gateway protocol
BI	busy indication
BIP	broadcasting incremental power
BP	aBeacon period
BPS	bounded-degree planar spanner
BPSK	binary phase shift keying
BSC	base station controller

CA	collision avoidance (CSMA/CA often)
CBC	cipher-block chaining (mode)
CBT	core-based tree
CBTC	cone-based topology control
CCA	clear channel assessment
CCM	combined cipher machine
CCR	critical coverage range
CDMA	code-division multiple-access
CDS	connected dominating set
CEDAR	core-extraction distributed ad hoc routing
CFB	cipher-feedback (mode)
CF-End	contention-free end
CFP	contention-free period
CF-Poll	contention-free poll
CG	conflict graph
CGSR	cluster-head gateway switch routing
CM	cross-monotone
CNN	critical neighbor number
CP	contention period
CPA	closest point of approach
CPU	central processing unit
CRC	cyclic redundancy check
CSMA	carrier-sense multiple-access (protocol)
CSP	collaborative signal processing
CT2	cordless telephone
CTR	critical transmission range
CTS	clear-to-send (mechanism)
CW	contention window
D/A	digital-to-analog (conversion)
DAG	directed acrylic graph
D-AMPS	digital advanced mobile phone service
DARPA	Defense Advanced Research Projects Agency
DC	differential cryptanalysis
DCA	dynamic channel assignment
DCF	distributed coordination function
DECT4	digital European cordless telephone
Demod	demodulator
DES	data encryption standard
DG	disk graph
D-H	Diffie–Hellman
DIFS	distributed interframe space
DM	dense model
D-PRMA	distributed packet-reservation multiple-access (protocol)
DPT	distributed prediction tracking
DRAND	a protocol that technically is defined as distributed randomized TDMA scheduling

DREAM	distance routing effect algorithm for mobility
DS	dominating set
DSA	digital-signature algorithm
DSDV	destination-sequenced distance-vector [routing (protocol)]
DSL	digital subscriber line
DSN	Distributed Sensor Networks (program)
DSP	digital signal processing
DSR	dynamic source routing
DSSS	direct-sequence spread spectrum
DST	directed Steiner tree
DT	Delaunay triangulation
DV	distance vector
DVMRP	distance-vector multicast routing protocol
EAX	designation of a two-pass authenticated encryption scheme
EC	Euler circuit
ECB	electronic codebook (mode)
ECC	elliptic curve cryptography
EDCA	enhanced DCF channel access
EDGE	enhanced data rate for GSM evolution
EFF	Electronic Frontier Foundation
EIFS	extended interframe space
ELSD	equal link split downstream
EMST	Euclidean minimum spanning tree
ERNG	extended relative neighborhood graph
ETX	expected transmission count
ExOR	name given to an opportunistic multihop routing protocol
FDMA	frequency-division multiple-access
FFT	fast Fourier transform
FGSS	fault-tolerant global spanning subgraph
FHSS	frequency-hopping spread spectrum
FIPS	Federal Information Processing Standard
FLSS	fault-tolerant local spanning subgraph
FM	frequency modulation
FNR	farthest-neighbor routing
FP	final permutation
fPrIM	fixed-protocol-interference model
FPTAS	fully polynomial-time-approximation scheme
FSK	frequency-shift-keying
GC	graph coloring
GFR	greedy–face routing
GG	Gabriel graph
GOAFR	greedy other adaptive face routing
GPRS	General Packet Radio Service
GPS	global positioning system

GPSR	greedy perimeter stateless routing
GRG	geometric random graph
GSM	Global System for Mobile Communication
GTFT	method proposed in a paper
HC	hybrid coordinator
HCCA	HCF-controlled channel access
HCF	hybrid coordination function
HRMA	hop reservation multiple access
IARP	intrazone routing protocol
IBSS	independent basic service set
IBSSID	IBSS identifier
IC	incentive compatible
ICDS	induced connected dominating set (graph)
ID	identification
IEEE	Institute of Electrical and Electronics Engineers
IF	intermediate-frequency
iff	if and only if
IG	interference graph
IMBM	iterative maximum-branch minimization
IMRG	incident MST and RNG graph
IMS	IP (Internet Protocol) Multimedia Subsystem
IP	integer programming (formulation)
IP	Internet protocol
IP	initial permutation
IPTV	Internet protocol television
IR	individual rationality
IS	independent set
ISM	Industrial, Scientific, and Medical
ISP	Internet service provider
IT	information technology
IV	initial value
IV	initialization vector
kbps	kilobits per second
kbytes	kilobytes
kNN	k-nearest-neighbor (classifier)
LAN	local-area network
LAR	location-aided routing
LBM	location-based multicast
LC	linear cryptanalysis
LCP	least-cost path
LCPT	least-cost path tree
LDEL	local Delauney graph
LEARN	localized-energy-aware restricted neighborhood (routing)
LLACK	link-layer acknowledgment

LMST	localized minimum spanning tree
LNA	low-noise amplifier
LP	linear programming
LPL	low-power listening
LSS	local spanning subgraph
LST	least-cost Steiner tree
MAC	medium-access control
MAN	metropolitan-area network
MANET	mobile ad hoc network
MAP	maximum a posteriori probability
MATSF	name of a protocol proposed in a paper (from MANET time synchronization)
MBGP	multiprotocol extension for a border gateway protocol
MBS	Mobile Broadband System
MC-CDMA	multicode CDMA
MCDS	minimum connected dominating set
MCG	mutual-communication graph
MCMT	minimum-cost multicast tree
MCU	microcontroller unit
MDS	minimum dominating set
MEMS	Micro-Electro-Mechanical Systems
MFR	most-forwarding routing
MG	mutual-inclusion graph
MGC	minimum graph-coloring (problem)
MIB	management information base
MIMO	multiple-input multiple-output
MIP	multicast independent protocol
MIS	maximum independent set
ML	maximum-likelihood (classifier)
MMAC	multichannel MAC
MNP	monotone nonincreasing property
Mod	modulator
MOSPF	multicast open shortest path first
MPR	multipoint relay
MSC	mobile switching center
MST	minimum spanning tree
MUP	multiradio unification protocol
MVC	minimum vertex cover
MWCDS	minimum weighted connected dominating set
MWIS	maximum weighted independent set
MWVC	maximum weighted vertex cover
NAV	network allocation vector
NFR	no-free-rider
NIC	network interface card
NIST	National Institute of Standards and Technology
NNG	nearest-neighbor graph

NNR	nonnegative sharing
NP	nondeterministic polynomial
NPH	NP-Hard
NST	node-weighted Steiner tree
OAFR	other adaptive face routing
OCB	offset codebook (mode)
OFB	output feedback (mode)
OFDM	orthogonal frequency-division multiplexing
OFSF	orthogonal fixed-spreading-factor (code)
OLSR	optimized link-state routing (protocol)
OSPF	open shortest path first
OURS	optimal unicast routing system
OVSF	orthogonal variable-spreading-factor (code)
PA	power amplifier
PACS	personal-access communications systems
PAN	personal-area network
P-BIP	pruned broadcasting incremental power
PC	point coordinator
PCF	point coordination function
PCI	peripheral component interconnect
PDA	personal digital assistant
PhIM	physical-interference model
PHY	physical-layer (specification)
PI	planar and internal-node
PIFS	point-coordination-function interframe space
PIM-SM	protocol-independent multicast-sparse mode
PKCS	Public Key Cryptography Standard
P-MST	pruned minimum spanning tree
POMDP	partially observable Markov decision process
PP	primal linear programming
PrIM	protocol-interference model
PRMA	packet-reservation multiple-access (scheme)
PS	power-save (state or mode)
PSM	power-saving mode
P-SPT	pruned shortest-path tree
PSTN	public-switched telephone network
PTAS	polynomial-time-approximation scheme
PTC	polynomial-time computability
PTDMA	probabilistic time-division multiple access
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quadrature phase-shift keying
RAD	random-assessment delay
RAM	random-access memory

RBOP	related neighborhood-graph-based broadcast-oriented protocol
RF	radio frequency
RIP	routing information protocol
RNG	relative neighborhood graph
RON	resilient overlay network
RP	rendevous point
RPB	reverse-path-broadcasting (scheme)
RPF	reverse-path forwarding (lookup)
RREP	route reply
RREQ	route request
RSA	Rivest–Shamir–Adleman
RSS	received signal strength
RTS	request-to-send (mechanism)
RWP	random-waypoint (model)
Rx	receive
SBT	share-based tree
SCADA	supervisory control and data acquisition
SCH	set-cover hard
SIFS	short interframe space
SINR	signal-to-interference-noise ratio
SIR	signal-to-interference ratio
SMS	short messaging service
SOP	spectrum opportunity
SPAN	a topology maintenance protocol proposed by Chen *et al*. (2002)
SPF	shortest path first
SPS	Standard Positioning Service
SPT	shortest-path tree
SSCH	slotted seeded channel hopping (protocol)
SSR	signal stability routing
SSR	security stochastic routing
STASF	a synchronization protocol proposed in a paper by Zhou and Lai (2005)
SURAN	Survivable Radio Network (project)
SVM	support vector machine
TA	trust authority
TACS	Total Access Communications System
TATSF	a synchronization protocol proposed in a paper
TBTT	target Beacon transmission time
TC	traffic class
TCP	transmission control protocol
TDM	time-division multiplexing
TDMA	time-division multiple-access
TDoA	time difference of arrival
ToA	time of arrival
TORA	temporarily ordered routing algorithm
TSF	timing synchronization function
Tx	transmit

TxIM	transmitter-interference model
TxoP	transmit opportunity
UDG	unit disk graph
UDP	user data-gram protocol
UMTS	Universal Mobile Telecommunication System
UPVCS	undirected minimum-power k-vertex-connected subgraph
US	ultrasound
UWB	ultrawideband
UWCDS	unicast weighted connected dominating set
VC	Vapnik and Chervonenkis
VCG	Vickrey–Clarke–Groves (mechanism)
VCO	voltage-controlled oscillator
VHF	very-high-frequency
VMST	virtual minimum spanning tree
VoIP	voice over IP
VOR	VHF omnidirectional ranging (aircraft navigation system)
VoWIP	voice over wireless IP
WAN	wireless ad hoc network
WCDMA	wideband code-division multiple-access
WCDS	weighted connected dominating set
WEP	wired equivalent privacy (encryption)
WiFi	common name used to refer to a wireless local-area network
WiMAX	Worldwide Interoperability for Microwave Access
WINS	a type of sensor node by Rockwell
WLAN	wireless local-area network
WMAN	wireless metropolitan-area network
WMN	wireless mesh network
WPA	WiFi protected access (mode)
WPAN	wireless personal-area network
WRP	wireless routing protocol
WSN	wireless sensor network
WWAN	wireless wide-area network
WWiSE	Worldwide Spectrum Efficiency (standard)
YG	Yao graph

Part I

Introduction

1 History of Wireless Networks

1.1 Introduction

The wireless arena has been experiencing exponential growth in the past decade. We have seen great advances in network infrastructures, rapid growth of cellular network users, the growing availability of wireless applications, and the emergence of omnipresent wireless devices such as portable or handheld computers, personal digital assistants (PDAs), and cellular phones, all becoming more powerful in their applications. The mobile devices are becoming smaller, cheaper, more convenient, and more powerful. They can also run more applications on the network services. For example, mobile users can rely on their cellular phones to check e-mail and browse the Internet. They can do so from airports, railway stations, cafes, and other public locations. Tourists can use the global positioning system (GPS) terminals installed in cars to view driving maps and locate attractions. All these factors are fueling the explosive growth of the cellular communication market. As of 2006, the number of cellular network users approached two billion worldwide. Market reports from independent sources show that worldwide cellular users have been doubling every 1.5 years.

In addition to that of the traditional cellular networks, an exponential growth of the wireless access point (AP), which is a device that connects wireless communication devices together to create a wireless network, is also being experienced. The AP is usually connected to a wired network and can relay data between devices on each side. Many APs can be connected together to create a larger network, which is a so-called *ad hoc network*. Low-cost, easily installed APs grew rapidly in popularity in the late 1990s and early 2000s. According to a new research study from Pyramid Research, WiFi users will outnumber cellular users by 2007. This trend will put increasing pressure on wireless operators to bundle both types of access. Currently, most of the connections among wireless devices occur over fixed-infrastructure-based service providers or private networks. Although the research and development efforts devoted to traditional wireless networks are still considerable, the interest of the scientific and industrial community of telecommunications has recently shifted to more challenging ad hoc wireless networks, in which a group of (potentially mobile) units equipped with radio transceivers can communicate without any fixed infrastructure. We will soon see a convergence of seamless networks that will keep everyone connected from their home to their office and all points in between. In addition, with the breakdown of traditional communications

infrastructures during the recent Hurricane Katrina catastrophe, the need for reliable connectivity in order for emergency responders to talk to each other is even greater.

1.2 Different Wireless Networks

A number of different wireless networks exist and can be categorized in various ways depending on the criteria chosen for their classification, such as network architecture and communication coverage area.

Based on Network Architecture

Wireless networks can be divided into two broad categories based on how the network is constructed, i.e., the underlying network architecture.

1. **Infrastructure-based networks:** An infrastructure-based network is a network that has a preconstructed infrastructure that is made of a fixed network structure (typically, wired network nodes and gateways). Network services are delivered via these pre-constructed infrastructures. For example, cellular networks are infrastructure-based networks, which are built from public-switched telephone network (PSTN) backbone switches, mobile switching centers (MSCs), base stations, and mobile hosts. Each node of the network has its specific responsibility in routing the data, and the connection establishment follows a strict signaling sequence among the nodes. Another example of infrastructure-based networks are wireless local-area networks (WLANs).
2. **Infrastructureless networks:** An infrastructureless network is a network that is formed dynamically through the cooperation of an arbitrary set of independent wireless devices. There is no prearrangement of the specific roles for each node. Typically, each node is assumed to be able to forward the data packets for any other node if it is asked to do so. Each node can independently make its own decision based on the network situation. Examples of infrastructureless wireless networks include mobile ad hoc networks and wireless sensor networks.

Another classification criterion for wireless networks is based on the communication coverage area of the networks.

Based on Communication Coverage Area

As with wired networks, wireless networks can be categorized into different types of networks based on the distances over which the data are transmitted.

1. **Wireless wide-area networks (WWANs):** WWANs are infrastructure-based networks that rely on networking infrastructures to enable mobile users to establish wireless connections over remote networks. These connections often could be over a very large geographic areas (across cities or even countries) through the use of multiple antenna sites or satellite systems maintained by wireless service providers. Examples of WWANs include cellular networks and satellite networks.

2. **Wireless metropolitan-area networks (WMANs):** WMANs are also infrastructure-based networks that enable users to establish broadband wireless connections among multiple locations within a metropolitan area without the high cost of laying fiber or copper cabling lines. Both radio waves and infrared light can be used in WMANs to transmit data. The U.S. Institute of Electrical and Electronics Engineers (IEEE) set up a specific 802.16 Working Group on Broadband Wireless Access Standards that develops standards and recommended practices to support the development and the deployment of WMANs.
3. **Wireless local-area networks (WLANs):** WLANs enable users to establish wireless connections within a local area, typically within a corporate or campus building or in a public space such as an airport. The connections are typically within a 100-m range. WLANs can operate in an infrastructure-based mode or in an infrastructureless mode. In the infrastructure-based mode, wireless stations connect to wireless APs that serve as bridges between the stations and an existing network backbone. In the infrastructureless mode, several wireless stations within a limited area form a temporary network without using the wireless APs if they do not require access to outside network resources. Typical examples of WLAN implementations include 802.11 (also called WiFi) and Hiperlan2.
4. **Wireless personal-area networks (WPANs):** WPAN technologies enable users to establish ad hoc wireless communication among personal wireless devices such as PDAs, cellular phones, or laptops that are within a personal operating space. A WPAN operates in infrastructureless mode, and the connections are typically within a 10-m range. Two key WPAN technologies are Bluetooth and infrared light. Bluetooth is a cable-replacement technology that uses radio waves to transmit data to a distance of up to 10 m, whereas infrared can connect devices within a range of 1 m. WPAN implementations often have low complexity, lower power consumption, and are interoperable with 802.11 networks.

1.2.1 Wireless Cellular Networks

First-Generation Mobile Systems

The first generation of analog cellular systems included the Advanced Mobile Telephone System (AMPS), which was made available in 1983. It was first deployed in Chicago, with a service area of 2100 square miles. AMPS offered 832 channels, with a data rate of 10 kilobits per second (kbps). Although omnidirectional antennas were used in the earlier AMPS implementation, it was realized that using directional antennas would yield better cell reuse. In fact, the smallest reuse factor that would fulfill the 18-db signal-to-interference and noise ratio (SINR) by use of 120-deg directional antennas was found to be 7. Hence, a 7-cell reuse pattern was adopted for the AMPS. Transmissions from the base stations to mobiles occur over the forward channel by use of frequencies between 869 and 894 MHz. The reverse channel is used for transmissions from mobiles to the base station, with frequencies between 824 and 849 MHz.

In Europe, the Total Access Communications System (TACS) was introduced with 1000 channels and a data rate of 8 kbps. AMPS and TACS use the frequency-modulation (FM) technique for radio transmission. Traffic is multiplexed onto a frequency-division multiple-access (FDMA) system. In Scandinavian countries, the Nordic Mobile Telephone is used.

Second-Generation Mobile Systems

Compared with first-generation systems, second-generation (2G) systems use digital multiple-access technology, such as time-division multiple access (TDMA) and code-division multiple access (CDMA). The Global System for Mobile Communications, or GSM, uses TDMA technology to support multiple users. Examples of 2G systems are the GSM, cordless telephone (CT2), personal-access communications systems (PACSs), and digital European cordless telephone (DECT4).

A new design was introduced into the MSC of 2G systems. In particular, the use of base station controllers (BSCs) lightens the load placed on the MSC found in first-generation systems. This design allows the interface between the MSC and the BSC to be standardized. Hence, considerable attention was devoted to interoperability and standardization in 2G systems so that a carrier could employ different manufacturers for the MSCs and BSCs. In addition to enhancements in MSC design, the mobile-assisted handoff mechanism was introduced. By sensing signals received from adjacent base stations, a mobile unit can trigger a handoff by performing explicit signaling with the network.

2G protocols use digital encoding and include the GSM, digital AMPs (D-AMPS) (TDMA), and CDMA (IS-95). 2G networks are in current use around the world. The protocols behind 2G networks support voice and some limited data communications, such as faxing and short messaging services (SMSs), and most 2G protocols offer different levels of encryption and security. Although first-generation systems support primarily voice traffic, 2G systems support voice, paging, data, and fax services.

2.5G Mobile Systems

The move into the 2.5G world began with the General Packet Radio Service (GPRS). GPRS is a radio technology for GSM networks that adds packet-switching protocols, a shorter setup time for Internet service provider (ISP) connections, and the possibility of charging by the amount of data sent rather than by connection time. Packet switching is a technique whereby the information (voice or data) to be sent is broken up into packets of, at most, a few kilobytes each, which are then routed by the network between different destinations based on addressing data within each packet. Use of network resources is optimized as the resources are needed only during the handling of each packet.

The next generation of data heading toward third-generation (3G) and personal multimedia environments builds on the GPRS and is known as the enhanced data rate for GSM evolution (EDGE). EDGE is also a significant contributor in 2.5G. It allows GSM operators to use existing GSM radio bands to offer wireless multimedia Internet-protocol-(IP-) based services and applications at theoretical maximum speeds of 384 kbps with

a bit rate of 48 kbps per time slot and up to 69.2 kbps per time slot in good radio conditions. EDGE will let operators function without a 3G license and compete with 3G networks offering similar data services. Implementing EDGE will be relatively painless and will require relatively small changes to network hardware and software because it uses the same TDMA frame structure, logic channel, and 200-kHz carrier bandwidth as today's GSM networks. As EDGE progresses to coexistence with 3G wideband CDMA (WCDMA), data rates of up to asynchronous-transfer-mode- (ATM-) like speeds of 2 Mbps could be available.

The GPRS will support flexible data transmission rates as well as a continuous connection to the network. The GPRS is the most significant step toward 3G.

Third-Generation Mobile Systems

3G mobile systems face several challenging technical issues, such as the provision of seamless services across both wired and wireless networks and universal mobility. In Europe, there are three evolving networks under investigation: Universal Mobile Telecommunications Systems (UMTSs), Mobile Broadband Systems (MBSs), and WLANs.

The use of hierarchical cell structures is proposed for IMT2000. The overlaying of cell structures allows different rates of mobility to be serviced and handled by different cells. Advanced multiple-access techniques are also being investigated, and two promising proposals have evolved, one based on WCDMA and another that uses a hybrid TDMA–CDMA–FDMA approach.

1.2.2 Wireless Access Points

A wireless AP is a device that connects wireless communication devices together to create a wireless network. The AP is usually connected to a wired network and can relay data between devices on each side. Many APs can be connected together to create a larger network that allows "roaming." In contrast, a network in which the client devices manage themselves is called an ad hoc network.

Low-cost, easily installed APs grew rapidly in popularity in the late 1990s and early 2000s. These devices offered a way to avoid tangled messes of cables associated with typical Ethernet networks of the day. Wireless networks also allowed users greater mobility, freeing individuals from the need to be stuck at a computer cabled to the wall. On the industrial and commercial side, wireless networking had a big impact on operations: Employees were often equipped with portable data terminals integrating bar-code scanners and wireless links, allowing them to update work-in-progress and inventory in real time.

One IEEE 802.11 AP can typically communicate with 30 client systems within a radius of 100 m. However, the communication range can vary a lot, depending on such variables as indoor or outdoor placement, height above ground, nearby obstructions, type of antenna, the current weather, operating radio frequency, and power output of the device. The range of APs can be extended through the use of repeaters and reflectors,

which can bounce or amplify radio signals that ordinarily could not be received. Some experiments have been carried out to allow wireless networking over distances of several kilometers.

A typical corporate use of an AP is to attach it to a wired network and then provide wireless client adapters for users who need them. Within the range of the AP, the wireless end-user has a full network connection with the benefit of mobility. In this instance, the AP is a gateway for clients to access the wired network. Another use is to bridge two wired networks for which cable is not appropriate; for example, a manufacturer can wirelessly connect a remote warehouse's wired network with a separate (though within line of sight) office's wired network.

An AP may also act as the network's arbitrator, negotiating when each nearby client device can transmit. However, in the vast majority of currently installed IEEE 802.11 networks, this is not the case, as a distributed pseudo-random algorithm is used instead.

Limitations

There are only a limited number of frequencies legally available for use by wireless networks. Usually, adjacent APs will use different frequencies to communicate with their clients in order to avoid interference between the two nearby systems. Wireless devices are able to "listen" for data traffic on other frequencies and can rapidly switch from one frequency to another to achieve better reception on a different AP. However, the limited number of frequencies becomes problematic in crowded downtown areas with tall buildings housing multiple APs because there can be enough overlap between the wireless networks to cause interference.

Wireless networking is far behind wired networking in terms of bandwidth and throughput. Whereas (as of 2007) typical wireless devices for the consumer market can reach speeds of 11 (IEEE 802.11b) or 54 megabits per second (Mbit/s) (IEEE 802.11a, IEEE 802.11g), wired hardware of similar cost reaches 1000 Mbit/s (Gigabit Ethernet). One impediment to increasing the speed of wireless communications is that WiFi uses a shared communications medium, so the actual usable data throughput of an AP is somewhat less than half the over-the-air rate. Thus, a typical 54-Mbit/s wireless connection actually carries TCP/IP (TCP stands for transmission control protocol) data at 20 to 25 Mbit/s. Because users of legacy wired networks are used to the faster speeds, people using wireless connections are anxious to see the wireless networks catch up.

Security

Another issue with wireless access in general is the need for security. Many early APs were not able to discern whether a particular user was authorized to access the network. Although this problem reflects issues that have long troubled many types of wired networks (it has been possible in the past for individuals to plug computers into randomly available Ethernet jacks and get access to the network), this was usually not a significant problem because many businesses had reasonably good physical security. However, the fact that radio signals bleed outside of buildings and across property lines means that physical security is not as much of a deterrent to war drivers.

In response, several new security technologies have emerged. One of the simplest techniques involves only allowing access from certain medium-access control (MAC) addresses. However, MAC addresses can be easily spoofed, leading to the need for more advanced security measures. Many APs incorporate a wired equivalent privacy (WEP) encryption, but that also has been criticized by many security analysts as not good enough (the U.S. FBI demonstrated the ability to break WEP protection in 3 min). Newer (as of 2005) encryption standards available on wireless APs and client cards include WiFi protected access, WPA and WPA2 modes, both of which offer substantial improvements in security. The WiFi alliance has announced the inclusion of additional Extensible Authentication Protocol (EAP) types to its certification program for WPA- and WAP2-Enterprise. Also, a newer system for authentication, IEEE 802.1x, promises to enhance security on both wired and wireless networks. Wireless APs that incorporate technologies like these often also have routers built in, so they are somewhat more accurately described as wireless gateways.

1.2.3 Wireless Ad Hoc Networks

A wireless ad hoc network is a collection of autonomous nodes or terminals that communicate with each other by forming a multihop radio network and maintaining connectivity in a decentralized manner. The wireless nodes communicate over wireless links; thus, they have to contend with the effects of radio communication, such as noise, fading, and interference. In addition, the links typically have less bandwidth than in a wired network. Each node in a wireless ad hoc network functions as both a host and a router, and the control of the network is distributed among the nodes. The network topology is in general dynamic, as the connectivity among the nodes may vary with time because of node departures, new node arrivals, and the change of environments. Hence, there is a need for efficient routing protocols to allow the nodes to communicate over multihop paths. Some of these features are characteristic of the type of packet radio networks that were studied extensively in the 1970s and 1980s. Recently, the wireless ad hoc networking research has received much attention from academia, industry, and government. Because these networks pose many complex issues, there are many open problems for research and opportunities for making significant contributions.

There are two major types of wireless ad hoc networks: mobile ad hoc networks and smart sensor networks.

Mobile Ad Hoc Networks

A mobile ad hoc network (MANET) is a self-configuring wireless network composed of wireless devices. Figure 1.1 illustrates an example of an ad hoc network formed by eight laptop computers. In the figure, two computers are connected by a line if they can communicate directly with each other by using their wireless cards. In this case, we say they are within the transmission range of each other. The wireless devices are free to move randomly and organize themselves arbitrarily. Consequently, the network topology may change rapidly and unpredictably. A MANET network may operate in a

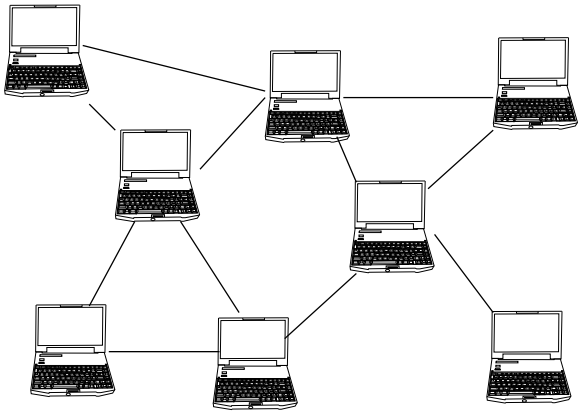


Figure 1.1 An ad Hoc network example.

stand-alone fashion or may be connected to the larger Internet. Because of their minimal configuration and quick deployment, ad hoc networks are often suitable for emergency situations like natural or human-induced disasters, military conflicts, emergency medical situations, and so on.

The earliest MANETs were called "packet-radio" networks, first sponsored by the U.S. Defense Advanced Research Projects Agency (DARPA) in the early 1970s. It is interesting to note that some early packet-radio systems predated the Internet and, indeed, were part of the motivation of the original Internet protocol (IP) suite. Later DARPA experiments included the Survivable Radio Network (SURAN) project, which took place in the 1980s. The third wave of academic activity on wireless ad hoc networks started in the 1990s, especially with the wide usage of inexpensive 802.11 radio cards for personal computers.

The popular IEEE 802.11 ("WiFi") wireless protocol incorporates an ad hoc networking system when no wireless APs are present. In an IEEE 802.11 system, each node transmits and receives data but does not route anything between the network's systems. Notice that it is possible to design higher-level protocols to aggregate various IEEE 802.11 ad hoc networks into MANETs.

Because of the growing interests in establishing survivable, efficient, dynamic communication for emergency/rescue operations, disaster relief efforts, and military networks, there is a strong need for the rapid deployment of independent mobile users. Obviously, we cannot rely on a centralized and organized network structure for these application scenarios. A MANET is an autonomous collection of mobile users that communicate over relatively bandwidth-constrained wireless links, for which all network activity including discovering the topology and delivering messages must be executed by the nodes themselves.

The design of network protocols for these networks is a complex issue. A unique characteristic of wireless networks is that the radio signal sent out by a wireless terminal will be received by all the terminals within its transmission range and also possibly causes signal interference to some terminals that are not intended receivers. In other words, the

communication channels are shared by the wireless terminals. Thus, one of the major problems facing wireless networks is the reduction of capacity because of interference caused by simultaneous transmissions. Using multiple channels and multiple radios can alleviate but not eliminate the interference. This raises the scalability issue of all wireless networks (MANETs, WSNs).

Regardless of the application, MANETs need efficient distributed algorithms to determine network organization, link scheduling, and routing. However, determining feasible routing paths and delivering messages in a decentralized environment where network topology fluctuates over time is not an easy problem, and, to some extent, it is even not a well-defined problem. Although the shortest path (based on a given cost function) from a source to a destination in a static wired network is usually the optimal route, this idea is not easily extended to MANETs. A number of unique characteristics of wireless networks make the "simple" optimal unicast routing much harder. For example, various factors, such as variable wireless link quality, propagation path loss, fading, multiuser interference, power expended, and topological changes, become relevant issues. Notice that finding the path (or even multiple paths) with the largest throughput to connect a given pair of source and target nodes in a wireless network is already a nondeterministic-polynomial- (NP-) hard problem even if only the wireless interference (interpath interference and intrapath interference) is considered. Moreover, in many applications such as a military environment, preservation of security, achieving small latency, reliability, preventing intentional jamming, and recovery from failure are significant concerns. This will make the design of a good wireless protocol much harder. Additionally, in certain applications (especially military networks), we need to maintain a low probability of intercept and/or a low probability of detection. Hence, nodes prefer to radiate as little power as necessary and transmit as infrequently as possible. A lapse in any of these requirements may degrade the performance and dependability of the network. Although there are so many challenges in designing secure and efficient wireless ad hoc networks, this book is not intended to (and, clearly, it is impossible to) solve all important and interesting problems here. Some of the algorithmic and graph theoretical issues that can form a foundation for further study of some of the problems not addressed here are covered.

Wireless Sensor Networks

Most sensors are electrical or electronic, although other types exist. A sensor is a type of transducer. Sensors are either direct indicating (e.g., a mercury thermometer or electrical meter) or are paired with an indicator [perhaps indirectly through an analog-to-digital (A/D) converter, a computer, and a display]. Sensors are heavily used, in addition to other applications, in medicine, industry, and robotics. With the technical progress, more and more sensors are manufactured with Micro-Electro-Mechanic-Systems (MEMS) technology. This often offers the potential of reaching a much higher sensitivity. A good sensor obeys the following rules:

1. The sensor should be sensitive to the measured property.
2. The sensor should be insensitive to any other property.
3. The sensor should not influence the measured property.

In the ideal situation, the output signal of a sensor is exactly proportional to the value of the measured property.

Distributed, wireless, microsensor networks will enable myriad applications for sensing and controlling the physical world. A WSN is a network made of hundreds or thousands of devices using sensors (also called nanocomputers) to monitor different conditions, such as temperature, sound, vibration, pressure, motion, or pollutants, at different locations. Usually these devices are small and inexpensive, so that they can be produced and deployed in large numbers. For example, for the field of computer science, most sensors are made by two companies, **xbow** and **moteiv**. One of the main differences between MANETs and WSNs is that the wireless sensors often have severely constrained resources in terms of energy, memory, computational speed, and bandwidth. The sensor nodes are self-contained units equipped with a radio transceiver, a small microcontroller, and an energy source, usually a battery. Recently, acoustic sensors have also been built for underwater monitoring. In most WNSs, the sensors typically rely on each other to transport data to a monitoring computer. The nodes dynamically self-organize their network topology based on various network conditions, rather than having a preprogrammed network topology. Because of the limitations that are due to battery life, nodes are built with power conservation in mind and generally spend large amounts of time in a low-power "sleep" mode or processing the sensor data. Thus, each sensor has wireless communication capability and some level of intelligence for signal processing and networking of the data. The wireless ad hoc sensor networks offer certain capabilities and enhancements in operational efficiency in civilian applications as well as in assisting in the national effort to increase alertness to potential terrorist threats. For almost all WSNs, there are three essential functions: sensing, communications, and computation (hardware, software, algorithms).

Modern research on sensor networks started around 1980 at DARPA: the Distributed Sensor Networks (DSN) program. Smaller computing chips, more capable sensors, wireless networks, and other new information technologies (ITs) are pushing the development of sensor networks.

There are many ways to classify the WSNs. One way is whether the nodes are individually addressable, and another is whether the data in the network are aggregated. Whether addressability is needed depends on the applications. For example, the sensor nodes in a parking-lot network should be individually addressable, so that one can determine the locations of all the free spaces. On the other hand, if one wants to determine the temperature in a corner of a room, then addressability may not be so important. The ability of the sensor network to aggregate the data collected can greatly reduce the number of messages that need to be transmitted across the network. In the majority of tasks of a WSN, an important requirement is that the required data be disseminated to the proper end-users. In some cases, there are fairly strict time requirements on this communication. For example, the detection of an intruder in a surveillance network should be immediately communicated to the police so that action can be taken.

To design an efficient and secure WSN, we often need to address a number of challenging issues.

Aside from the deployment of sensors on the ocean surface or the use of mobile, unmanned, robotic sensors in military operations, most nodes in a smart sensor network are stationary. Networks of 10,000 or even 100,000 nodes are envisioned, so scalability is a major issue. The algorithms and protocols designed for WSNs should bound the communication cost with respect to the network size.

Because in many applications the sensor nodes are often powered by batteries and will be placed in a remote area, recharging the batteries of a node may not be possible. In this case, the lifetime of a node may be determined by the battery life. Consequently, the minimization of energy expenditure is crucial. There are a considerable number of research papers in the literature that propose reducing the energy consumption of WSNs by use of various approaches such as topology control, data aggregation, data compression, energy-efficient MAC protocols, and smart use of some of the properties of batteries (a battery will last longer if it is not used continuously for a long time).

Given the large number of nodes and their potential placement in hostile locations, it is essential that the network be able to self-organize. Moreover, nodes may fail (either from lack of energy or from physical destruction and so on), and new nodes may join the network. Therefore, the network must be able to periodically reconfigure itself so that it can continue to function. Individual nodes may become disconnected from the rest of the network, but a certain connectivity of the network (or large portion of the network) must be maintained.

Another unique factor that distinguishes WSNs from MANETs is that the end goal is the detection and estimation of some events of interest and not just communications. To improve the detection–estimation performance, it is often quite useful to fuse data from multiple sensors. This data fusion requires the transmission of data and control messages, and so it may put constraints on the network architecture. Another important phenomenon is that we need to be able to distinguish between false data collected and the data reflecting a real emergency (e.g., a fire) in certain area. For example, a high temperature in an area reported by a sensor may indicate that there is a fire or may be due to errors in sensing or processing.

All WSNs should provide querying ability. A user may want to query an individual node or a group of nodes for information collected in the region. Depending on the amount of data fusion performed, it may not be feasible to transmit a large amount of the data across the network. Instead, various local sink nodes will collect the data from a given area and create summary messages. A query may be directed to the sink node nearest to the desired location.

The last, but not least, important challenge is the interoperability issue. With the coming availability of low-cost, short-range radios, along with advances in wireless networking, it is expected that WSNs will become commonly deployed. In these networks, each node may be equipped with a variety of sensors, such as acoustic, seismic, infrared, and still/motion videocamera. These nodes may be organized in clusters and coordinate with each other such that a locally occurring event can be detected by most if not all of the nodes in a cluster. These nodes will collaborate to make certain local decisions based on the information and decisions collected from each individual node

within the cluster. One node may act as the cluster master, and it will coordinate these efforts.

Mesh Networks

Wireless mesh networking is mesh networking implemented over a WLAN. This type of Internet infrastructure is decentralized, relatively inexpensive, and very reliable and resilient because each node need transmit only as far as the next node. Nodes act as repeaters to transmit data from nearby nodes to peers that are too far away to reach, resulting in a network that can span large distances, especially over rough or difficult terrain. Extra capacity can be installed by the addition of more nodes or the use of more channels. Mesh networks may involve either fixed or mobile devices. Wireless mesh networks (WMNs) are being used as the last mile for extending the Internet connectivity for mobile nodes. Many U.S. cities (e.g., Medford, Oregon; Chaska, Minnesota; and Gilbert, Arizona) have already deployed mesh networks. Accesos Web Alternativos (AWA), the Spanish operator of WLAN networks, will roll out commercial WLANs and WMNs for voice and data services. Several companies such as MeshDynamics have recently announced the availability of multihop multiradio mesh network technology. These networks behave almost like wired networks because they have infrequent topology changes, limited node failures, and so on. For WMNs, the aggregate traffic load of each routing node also changes infrequently.

The choice of radio technology for WMNs is crucial. In a traditional wireless network in which laptops connect to a single AP, the more laptops connected, the less bandwidth available for each user. This is because the devices share a fixed bandwidth amount. With mesh technology and adaptive radio, devices in a mesh network will connect only with other devices that are in a set range. The advantage is that like a natural load-balancing system, the more devices, the more bandwidth that becomes available, provided that the number of hops in the average communications path is kept low. To prevent an increased hop count from canceling out the advantages of multiple transceivers, one common type of architecture for a mobile mesh network includes multiple fixed-base stations that will provide gateways to services, wired parts of the Internet, and other fixed-base stations.

1.3 Conclusion

Many people were involved in the invention of radio transmission of information as we know it today. Despite this, during its early development and long after wide acceptance, disputes persisted as to who could claim sole credit for this invention. James Clerk Maxwell performed the theoretical physical research that correctly predicted the existance of radio (and all other electromagnetic) waves. David E. Hughes was the first to transmit Morse code by radio, but scientists of his time were not quick to recognize Maxwell's theories nor Hughes' experiments. Heinrich Rudolf Hertz was the experimental physicist who first created radio waves in a controlled manner. Later

developments are greater or lesser engineering developments of their work. In late 1896 or early 1897, Nikola Tesla (10 July 1856–7 January 1943) received wireless signals transmitted from the Houston Street lab in New York City to West Point. Marconi began to conduct experiments, building much of his own equipment in the attic of his home at the Villa Griffone in Pontecchio, Italy. Marconi transmitted radio signals for about a mile at the end of 1895. In 1904, Marconi got his own patent, declaring principles that Tesla had developed. The issue of patent infringement by Marconi was addressed in a lawsuit brought by Tesla in 1915. In 1943, the Supreme Court of the United States credited Nikola Tesla as being the inventor of the radio. The first radio telephone network for commercial use was made available to consumers by the Bell Telephone Company in the early 1950s. In 1971, the world's first WLAN, named ALOHAnet, was developed by researchers at the University of Hawaii. These days, the use of GSM, PCS, and WiFi has spread to almost every corner in the world. In the past 10 years, wireless ad hoc networks and recently WSNs have drawn a considerable amount of research interests from various research fields. WSNs build a connection between the physical world that still has many unknowns left for exploring and the digital world dominated by the Internet. It is expected that WSNs will dramatically improve our understanding in many fields. In this chapter, only a brief review of the development history of wireless networks was given and some categorization of wireless networks was presented. A more detailed review on this topic can be found in many great books and on Internet websites.

Problems

1.1 What are the major differences between wired networks and wireless networks? Why do some problems that are easy to solve for wired networks becomes NP-hard for wireless networks? Can you list a few such questions?

1.2 In what band do most cellular and WLAN systems operate?

1.3 Find the relationship among bandwidth, information capacity, and the signal-to-interference-noise ratio.

1.4 What are the advantages and disadvantages of WiFi?

1.5 Compare and contrast the advantages and disadvantages of communicating via MANETs and via cellular networks.

1.6 What are the advantages and disadvantages of operating in unlicensed bands?

1.7 What are the advantages and disadvantages of operating in licensed bands?

1.8 What are the differences among MANETs, WLANs, WSNs, and WMNs?

1.9 Find and read good survey articles about MAC protocols, routing protocols, and mobility management protocols for MANETs.

1.10 Find and read good survey articles about MAC protocols, routing protocols, localization protocols, data processing, and aggregation protocols, and energy-efficient topology control protocols for WSNs.

1.11 Find and read good survey articles about MAC protocols, routing protocols, and cross-layer design protocols for WMNs.

1.12 What are the next-generation wireless networks? What is opportunistic spectrum usage? What is the current status? What are the major challenges in networks with opportunistic spectrum usage?

2 Wireless Transmission Fundamentals

In this chapter, some simple but also widely accepted models of wireless ad hoc networks are introduced. Notice that WSNs comprise a special subclass of wireless ad hoc networks; thus, when we use the term "wireless ad hoc networks," we also include WSNs if not specifically clarified. In this chapter, the main focus is on the wireless channel model, the interference model, the energy-consumption model, and the mobility model.

2.1 Wireless Channels

The main difference between wireless networks and traditional wired networks is that the wireless devices in a network communicate over wireless channels via wireless transceivers. Thus, to understand wireless ad hoc networks and design efficient protocols and algorithms for wireless networks, we need to understand the characteristics of wireless communications. An important building block of wireless ad hoc network studies is thus the wireless channel model. In the literature, there are a number of wireless channel models proposed and the model presented in this chapter is based on the material contained in Rappaport (1996) and Santi (2005b).

It is widely assumed that a radio channel from a transmitting wireless device u to a receiving wireless device v is established if and only if the received power of the radio signal at node v is above a certain threshold. Let $\mathbf{p}(u, v)$ denote the power assigned to node[1] u to transmit a signal from u to v. We always assume that this power can maintain a reasonably good communication link quality[2] from node u to node v. This power $\mathbf{p}(u, v)$ could be fixed throughout the network operations if no power-control techniques are employed, or it could be changed dynamically when it is needed by the power-control techniques or to ensure energy-efficient routing. It is well known that wireless propagation suffers from severe attenuation. Let $\|uv\|$ denote the Euclidean distance between two wireless nodes u and v. If node u transmits at a power $P_t(u)$, the power of the signal received at node v is assumed to be

$$P_r(v) = \frac{P_t(u)}{g(u, v)},$$

[1] In this book, the term *node* often represents a network device, *vertex* is a graph term, and *point* is a geometry term. We often interchange them if no confusion is caused.

[2] In practice, it often means that the link error probability is not larger than a certain threshold.

where $g(u, v)$ is the wireless gain between node u and v [$1/g(u, v)$ is often called *path loss* in the literature]. It is commonly assumed in the literature that we can always correctly decode the signal when the received power $P_r(v)$ satisfies $P_r(v) \geq \beta_0 N_0$, where β_0 is the required minimum *signal-to-interference-noise ratio* (SINR) and N_0 is the strength of the ambient noise. Here, the constant β_0 is wireless technology and device dependent. Thus, by assuming that the node u transmits at power $P_r(u) \geq \beta_0 N_0 g(u, v)$, it is assumed in the literature that we can guarantee that node v will receive the signal correctly.

Notice that, in practice, this is not the case. When a node u transmits at a power p to another node v, the link (u, v) has a packet-error probability dependent on the transmission power p. Notice that the packet-error probability also depends on other factors, such as the environment, the digital modulation techniques, and so on.

Modeling the path loss has historically been one of the most difficult tasks of the wireless-system designer because it depends on many parameters such as location, time, weather, and so on. A radio propagation model is an empirical mathematical formulation for the characterization of radio-wave propagation as a function of frequency, distance, and other conditions. A single model is usually developed to predict the behavior of propagation for all similar links under similar constraints. In the remainder of this section, we review some of the widely used radio propagation models used in the literature.

2.1.1 Free-Space Propagation Model

The free-space propagation model assumes the ideal propagation condition that there is only one clear line-of-sight path between the transmitter and receiver. In other words, this model can be used to predict radio-signal propagation when the path between the transmitter and the receiver is clear and without obstruction. Let P_t be the power used by the transmitter to transmit the radio signal. Let $P_r(d)$ be the power of the radio signal received by a node located at distance d from the transmitter. Under the free-space propagation model, we have

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L}, \tag{2.1}$$

where G_t is the transmitter antenna gain, G_r is the receiver antenna gain, $L \geq 1$ is the system loss factor independent of the propagation, and λ is the wavelength in meters. By ignoring the specific characteristics of the transmitter and the receiver, we can simplify Eq. (2.1) as follows:

$$P_r(d) = \frac{C_f P_t}{d^2}, \tag{2.2}$$

where C_f is a constant depending on the characteristics of the transmitter and the receiver. Here, f stands for *free space*. In certain scenarios (e.g., the devices are uniform), it is often assumed in the literature that $C_f = 1$. Notice that it is common to select $G_t = G_r = 1$ and $L = 1$ in NS_2 simulations.

When the sensitivity of the receiver node is given as β_0 and the background noise and interference are denoted as N_0, we can state that the transmitted message can be correctly received if and only if the distance d satisfies

$$d \leq \sqrt{\frac{C_f P_t}{\beta_0 N_0}}.$$

In other words, the free-space model basically represents the communication range as a disk (or sphere in three dimensions) around the transmitter with radius $\sqrt{\frac{C_f P_t}{\beta_0 N_0}}$. If a receiver is within the disk, it receives all packets. Otherwise, it loses all packets.

In practice, the free-space propagation model is valid only for values of d that are relatively far from the transmitting antenna. For values of d within the so-called close-in distance d_0, the path loss can be assumed to be constant.

2.1.2 Two-Ray Ground Model

For the free-space propagation model, it assumes that the single direct path between the transmitter and the receiver is the only physical means of propagation of the radio signal. In practice, it is rarely the case and thus the free-space model is often inaccurate, although it is widely adopted in the literature. The two-ray ground-reflection model considers both the direct path and a ground-reflection path. It is shown in Rappaport (1996) that this model gives a more accurate prediction at a long distance than the free-space model. The received power at distance d is predicted by

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4 L}, \tag{2.3}$$

where h_t and h_r are the heights of the transmitting and receiving antennas, respectively. Notice that in some literature [e.g., Rappaport (1996)], it is assumed that $L = 1$. To be consistent with the free-space propagation model, some works also add L here.

Equation (2.3) shows a faster power loss than the model for the free-space propagation model [Eq. (2.1)] as distance increases. However, the two-ray model does not give a good result for a short distance because of the oscillation caused by the constructive and destructive combination of the two rays. Instead, the free-space model is still used when d is small. When the distance d between the transmitter and the receiver is relatively large ($d \gg \sqrt{h_t h_r}$), we can abstract the features of the radio transmitters and receivers and get the following simplified formula:

$$P_r(d) = \frac{C_t P_t}{d^4}, \tag{2.4}$$

where C_t is a constant depending on the characteristics of the transmitter and the receiver. Here, t stands for *two-ray ground*.

When the sensitivity of the receiver node is given as β_0 and the background noise and interference are denoted as N_0, we can state that the transmitted message can be

correctly received if and only if the distance d satisfies

$$d \leq \sqrt[4]{\frac{C_t P_t}{\beta_0 N_0}}.$$

In other words, the free-space model basically represents the communication range as a circle (or sphere in three dimensions) around the transmitter with radius $\sqrt[4]{\frac{C_t P_t}{\beta_0 N_0}}$. If a receiver is within the circle, it receives all packets. Otherwise, it loses all packets.

Obviously, the main difference between the free-space propagation model and the two-ray ground model is that the signal falloff is proportional to the distance raised to the fourth power in the two-ray ground model here, whereas it is the distance raised to the square. Therefore, a crossover distance d_c is computed in these two models. When $d < d_c$, the free-space-propagation model is used, and the two-ray ground model is used otherwise. At the crossover distance d_c, these two models should give the same result. Thus, we can compute d_c as

$$\frac{4\pi h_t h_r}{\lambda}.$$

2.1.3 The Log-Distance Path-Loss Model

The log-distance model is derived from the combination of analytical and empirical methods. The log-distance path-loss model is an indoor radio-propagation model that predicts the path loss a signal encounters inside a building over a distance. The log-distance path-loss model is formally expressed as

$$L = L_0 + N \log \frac{d}{d_0} + X_g,$$

where L is the total path loss inside a building [with units of decibels (dB)], L_0 is the path-loss at reference distance (usually, 1 km or 1 mile with units of dB), N is the path-loss distance exponent times 10, and X_g is a Gaussian random variable with zero mean and σ standard deviation. The coefficients N and σ depend on the environment and also on the frequency of the radio.

The log-distance model can be seen as a generalization of both the free-space and the two-ray-ground propagation model. In other words, the average long-distance path loss is proportional to the separation distance d raised to a certain exponent α, which is called the *path-loss exponent* or *distance-power gradient*. In most literature, it is assumed that

$$P_r(d) \propto \frac{P_t}{d^\alpha}.$$

In other words, the log-distance model also represents the communication range as a disk (or sphere in three dimensions) around the transmitter with radius $\sqrt[\alpha]{\frac{C_f P_t}{\beta_0 N_0}}$. If a receiver is within the disk, it receives all packets. Otherwise, it loses all packets. The exact value of α depends on the environmental conditions, and it has been evaluated in many scenarios. See Rappaport (1996) for more information about the empirical values of α.

2.1.4 Large-Scale and Small-Scale Variations

Notice that all the previously discussed propagation models predict the *average* received power at a certain distance from the transmitter. In practice, the intensity of the received signal is often denoted by a random variable, and its actual value can vary a lot from the predicted average value. Thus, probabilistic models have been used to account for this time- and location-dependent wireless channel. In a probabilistic propagation model, the coverage region of a radio (i.e., the region where a receiver can get the signal correctly) is no longer a disk. Two different classes of probabilistic propagation models have been discussed in Rappaport (1996). One of the models is the *large-scale model*, which predicts the variations of the signal intensity over large distances. The other one is the *small-scale model*, which predicts the variations of the signal intensity over very short distances. They are also called *multipath-fading models*. The shadowing model extends the ideal circle model to a richer statistic model: Nodes can probabilistically communicate only when near the edge of the communication range.

One of the most important large-scale models is the shadowing model, in which the path loss at distance d is modeled as a random variable with log-normal distribution centered about the mean value. The most important fading model is the Rayleigh model, which models small-scale variations of the signal intensity according to a random variable with Rayleigh distribution. See Rappaport (1996) for a more detailed discussion of the radio propagation models.

Observe that accounting for large-scale and small-scale variations of the radio signal is very complicated and renders the link model tightly coupled with the specific application environment. Thus, in this section and the remainder of the book, we model the wireless channel by using the log-distance path-loss model, which already extracts a considerable number of characteristics of the environment. This assumption is often a standard in conducting the theoretical and algorithmic study of wireless ad hoc networks (especially in the area of the topology control, power assignment, and so on).

2.2 The Wireless Communication Graph

The communication graph defines the network topology formed by a set of wireless devices; that is, the set of wireless links that these wireless devices can use to communicate with each other, possibly using multihop paths. Based on the discussion of the radio-signal propagation models in the previous section, obviously whether two devices (called nodes also hereafter) u and v can form a communication link (u, v) depends on (1) the relative Euclidean distance between these two nodes, (2) the transmitting power used by the transmitter to send the signal, and (3) the surrounding environment, which will determine which propagation model can be used.

Assume that there is a set $V = \{v_1, \ldots, v_n\}$ of n communication wireless terminals deployed in a region Ω (which could be some area in a two-dimensional plane or a region in a three-dimensional space). We also abuse the notation a little by using v_i to denote not only the identification of a wireless node but also the geometry position of this node.

If nodes are mobile, the physical node location is time-dependent. We always assume that the nodes will move within the original deployed region Ω. For simplicity, we use $v_i(t)$ to denote the location of node v_i at time t.

The complete communication graph is a *directed* graph $G = (V, E)$, where $V = \{v_1, \ldots, v_n\}$ is the set of terminals and E is the set of possible *directed* communication links between pairs of wireless terminals. Let $E^-(u)$ denote the set of directed links that end at node u [i.e., (w, u)] and let $E^+(u)$ denote the set of directed links that start at node u [i.e., (u, v)].

Every terminal v_i has a transmission range $R_T(v_i)$ such that the necessary condition for a terminal v_j to receive correctly the signal from v_i is $\|v_i - v_j\| \leq R_T(v_i)$, where $\|v_i - v_j\|$ is the Euclidean distance between v_i and v_j. In other words, the transmission range of a node v_i denotes the maximum distance within which the data transmitted by node v_i can be correctly received by the receiver. Given the transmission range r of a node v_i, the definition of the transmission region depends on the dimension of the network deploy region. In the case of one-dimensional networks, the transmission region is simply the segment of length $2r$ centered at node v_i; in the case of two-dimensional networks, the transmission region is simply the disk of radius r centered at node v_i; in the case of three-dimensional networks, the transmission region is simply the sphere of radius r centered at node v_i. Notice that here the transmission range of a node v_i is dependent on the transmission power of this node and the propagation model used. Throughout this book, we always assume that all nodes adopt the same signal-propagation model. Thus, the transmission range of any node is uniquely determined by its transmission power, and sometimes they are interchangeably used in the rest of this book. In most of the results presented in this book, we assume that $\|v_i - v_j\| \leq R_T(v_i)$ is *also* the sufficient condition for $(v_i, v_j) \in E$.

When all nodes have the same transmission range, then the resulting communication graph is often called a *unit disk graph* (UDG). In other words, we normalize the transmission range of each node to *one unit* and, consequently, two nodes can communicate with each other directly if and only if their distance is no more than one unit.

When different nodes may have different transmission powers (and thus different transmission ranges), several different models of communication graphs are used in the literature. One model requires that only undirected links be used to support communication. Thus, there is an *undirected link* $v_i v_j$ in the communication graph G if and only if node v_j is inside the transmission region of node v_i and v_i is inside the transmission region of node v_j. In other words, both directed links (v_i, v_j) and (v_j, v_i) exist. This model is called the *mutual-inclusion* graph model in X.-Y. Li *et al.* (2005c). Mathematically, a link $v_i v_j$ is included in the communication graph G if and only if the Euclidean distance $\|v_i - v_j\| \leq \min[R_T(v_i), R_T(v_j)]$. Another model will use all directed links for communication. Thus, when node v_j is inside the transmission region of node v_i, the communication graph will include a *directed* link (v_i, v_j).

Notice that, in practice, $\|v_i - v_j\| \leq R_T(v_i)$ is *not* the sufficient condition for $(v_i, v_j) \in E$. For some results presented in this book, we assume the latter case. Some links do not belong to G because of either the physical barriers or the selection of routing protocols. This model has been used to study various problems in the literature,

e.g., Kumar *et al.* (2005) and W. Wang *et al.* (2006b). We always use (v_i, v_j) to denote the *directed* link (v_i, v_j) hereafter if the communication graph is assumed to have directed links; i.e., node v_j is inside the transmission region of node v_i. We simply use $v_i v_j$ to denote an undirected link between two nodes; i.e., node v_i can directly receive the signal correctly from v_j and vice versa. When this is the case, we call the network a *general geometry graph*.

2.3 Power Assignment and Topology Control

A wireless node can receive the signal from another node if it is within the transmission region of the sender. Otherwise, they communicate through multihop wireless links by using intermediate nodes to relay the message. A larger transmission range of a wireless node means that more neighbors can communicate directly, but it costs more energy. Energy conservation is a critical issue in a wireless ad hoc network for the individual node and the network because the nodes are powered by small batteries only. For example, in a battlefield scenario, soldiers may not have time to replace or recharge the batteries of their wireless devices, and running out of batteries means a loss of all of their communication capacity. Thus, a considerable amount of research efforts focus on designing minimum-power-assignment algorithms to save energy for typical network tasks such as broadcast transmission (Clementi *et al.*, 2001b; Huiban and Verhoeven, 2004; Wan *et al.*, 2002b; Wieselthier *et al.*, 2000), routing (Srinivas and Modiano, 2003), connectivity (Althaus *et al.*, 2003; Blough *et al.*, 2002; Chen and Huang, 1989; Clementi *et al.*, 2000c; Kirousis *et al.*, 2000), and fault tolerance (Călinescu and Wan, 2003; Cheriyan *et al.*, 2002; Hajiaghayi *et al.*, 2003).

Power Assignment

We assume that the power w_{uv} needed to support the communication between two nodes u and v is a monotone increasing function of the Euclidean distance $\|uv\|$. In other words, $w_{uv} > w_{xy}$ if $\|uv\| > \|xy\|$ and $w_{uv} = w_{xy}$ if $\|uv\| = \|xy\|$. For example, in the literature it is often assumed that $w_{uv} = c + \|uv\|^{\alpha}$, where c is a positive constant real number, and real number $\alpha \in [2, 5]$ depends on the transmission environment. We also assume that all nodes have omnidirectional antennas; i.e., if the signal transmitted by a node u can be received by a node v, then it will be received by all nodes x with $\|ux\| \leq \|uv\|$. In addition, all nodes can adjust the transmission power dynamically. Specifically, each node u has a maximum transmission power $\mathbf{P}_{\max}$ and it can adjust its power to be exactly w_{uv} to support the communication to another node v. Consequently, if all wireless nodes transmit in their maximum power, they define a network that has a link uv iff $w_{uv} \leq \mathbf{P}_{\max}$. When nodes adjust their power dynamically, we say that a node u can reach a node v in an *asymmetric* communication model if node u transmits at a power of at least w_{uv}. Notice that here, in asymmetric communications, node v may transmit at a power less than w_{vu} and thus cannot reach u. We say that a node u can reach a node v in a *symmetric* communication model if both nodes u and v transmit at a power of at least w_{uv}.

An observation of power adaption is that the network topology is entirely dependent on the transmission range of each individual node. Links can be added or removed when a node adjusts its transmission range. A *power assignment* $\mathcal{P}$ is an assignment of power setting $\mathcal{P}(v_i)$ to wireless node v_i. Given a power assignment $\mathcal{P}$, we can define an induced direct communication graph $\overrightarrow{G}_{\mathcal{P}}$ in which there is a directed edge $\overrightarrow{uv}$ if and only if $w_{uv} \leq \mathcal{P}(u)$. We define the induced undirected communication graph G_P in which there is an edge uv if and only if $w_{uv} \leq \mathcal{P}(u)$ and $w_{vu} \leq \mathcal{P}(v)$. We hereby refer to $G_{\mathcal{P}}$ as the *induced communication graph* by power assignment $\mathcal{P}$. If all wireless nodes transmit at their maximum power $\mathbf{P}_{\max}$, the induced communication graph is called the *original communication graph* (UDG), which provides information about all possible topologies in accordance with characteristics of the wireless environment and node power constraints. In other words, all possible achievable network topologies are subgraphs of the original communication graph.

On the other hand, given a subgraph $G = (V, E)$ of the original communication graph, we can also extract a minimum power assignment $\mathcal{P}_G$, where

$$\mathcal{P}_G(u) = \max_{\{v|uv \in E\}} w_{uv},$$

to support the subgraph. We call this $\mathcal{P}_G$ an *induced power assignment* from G.

There are a number of optimization criteria studied in the literature for power assignment. The *min–max power-range assignment* is to find a power assignment whose maximum power is minimized among all possible power assignments that can achieve a certain network property (e.g., the induced communication graph is connected). The *min–total power-range assignment* is to find a power assignment whose total assigned power to all nodes is minimized among all possible power assignments that can achieve a certain network property (e.g., the induced communication graph is connected). A number of network properties have been studied in the literature for power assignment such as connectivity, two-connectivity, and generally k-connectivity. Both centralized and distributed (or even localized) algorithms have been proposed.

Because of the importance of energy efficiency in wireless ad hoc networks, minimum power assignments for different network issues have been addressed recently. Research efforts have focused on finding the minimum power assignment so that the induced communication graph has some "good" properties in terms of network tasks such as disjoint paths, connectivity, or fault tolerance. The minimum-energy-connectivity problem was first studied by Chen and Huang (1989), in which the induced communication graph is strongly connected while the total power assignment is minimized. They showed that this problem is NP-hard. Recently, this problem was intensively studied, and many approximation algorithms were proposed when the network is modeled by use of symmetric links or asymmetric links (Althaus *et al.*, 2003; Blough *et al.*, 2002; Călinescu *et al.*, 2003; Clementi *et al.*, 2000c; Kirousis *et al.*, 2000; Ramanathan and Rosales-Hain, 2000). Along this line, several authors (Călinescu and Wan, 2003; Cheriyan *et al.*, 2002; Hajiaghayi *et al.*, 2003) considered the minimum total power assignment while the resulting network is k-strongly connected or k-connected. This problem has been shown to be NP-hard too. Solving this problem can improve the fault

tolerance of the network. Clementi *et al.* (2000a, 2000b, 2000c) also considered the minimum-energy-connectivity problem while the induced communication graph has a diameter bounded by a constant h. Lloyd *et al.* (2002) proposed one general framework that leads to an approximation algorithm for minimizing total power assignment. Using the framework, they proposed a new two-connected approximation method for min–total power assignment. Krumke *et al.* (2003) also studied the minimum power assignment so that networks satisfy specific properties such as connectivity, bounded diameter, and minimum node degree. Other relevant work in the area of power assignment (also called energy efficiency) includes energy-efficient broadcasting and multicasting in wireless networks. The problem, given a source node s, is to find a minimum power assignment such that the induced communication graph contains a spanning tree rooted at s. This problem was proved to be NP-hard. In Clementi *et al.* (2001a), Huiban and Verhoeven (2004), Wan *et al.* (2002b), and Wieselthier *et al.* (2000), the authors presented some heuristic solutions and gave some theoretical analysis. Recently, Srinivas and Modiano (2003) also studied finding k-disjoint paths for a *given* pair of nodes while minimizing the total node power needed by nodes on these k-disjoint paths. An excellent survey of some recent theoretical advances and open problems on energy consumption in ad hoc networks can be found in Clementi *et al.* (2002).

Asymptotic Power Assignment

The previous discussion on power assignment assumed that there is *one fixed* network as the input. In the literature, there are a number of studies on power assignment that assume that the networking nodes could be mobile or there are a set of network instances as the input (whose number could be infinity). When the network nodes are mobile, a range assignment $\mathcal{P}$ is said to be *connected* at time t if the induced communication graph at time t is connected. Notice that the power assignment $\mathcal{P}$ could be a function of time t.

The universal minimum power used by all wireless nodes such that the induced network topology is connected is called the *critical power*. Determination of the critical power in which the wireless nodes are statically distributed was studied by several researchers recently (Gupta and Kumar, 1998; Ramanathan and Rosales-Hain, 2000; Sanchez *et al.*, 1999). Both Ramanathan and Rosales-Hain (2000) and Sanchez *et al.* (1999) use the power assignment induced by the longest incident edge of the Euclidean minimum spanning tree over wireless nodes V. Although determining the critical power for static wireless ad hoc networks is well studied, it remains to study the critical power for connectivity for mobile wireless networks. Because the wireless nodes move around, it is impossible to have a unanimous critical power to guarantee the connectivity for all instances of the network configuration. Thus, we need to find a critical power, if possible, at which each node has to transmit to guarantee the connectivity of the network almost surely (i.e., with a high probability sufficiently close to 1). For simplicity, we assume that the wireless devices are distributed in a unit square (or disk) according to some distribution function (e.g., uniform distribution or Poisson process). Additionally, we assume that the movement of wireless devices still keeps them the same distribution

(uniform or Poisson process). Gupta and Kumar (1998) showed that there is a critical power almost surely when the wireless nodes are randomly and uniformly distributed in a unit area disk. The result by Penrose (1998) implies the same conclusion. Moreover, Penrose (1998) gave the probability of the network's being connected if the transmission radius is set as a positive real number r and the number of nodes n goes to infinity.

Let $G(V, r)$ be the graph defined on V with edges $uv \in E$ if and only if $\|uv\| \leq r$. Here, $\|uv\|$ is the Euclidean distance between nodes u and v. Let $\mathcal{G}_\Omega(\mathcal{X}_n, r_n)$ be the set of graphs $G(V, r_n)$ for n nodes V that are uniformly and independently distributed in a two-dimensional region Ω, which could be a unit-area disk $\mathcal{D}$ or a unit square $\mathcal{C}$ with its center at the origin. The problem considered by Gupta and Kumar (1998) is then to determine the value of r_n such that a random graph in $\mathcal{G}_\mathcal{D}(\mathcal{X}_n, r_n)$ is asymptotically connected with probability one as n goes to infinity. Let $P_{\Omega,k}(\mathcal{X}_n, r_n)$ be the probability that a graph in $\mathcal{G}_\Omega(\mathcal{X}_n, r_n)$ is k-connected. Then Gupta and Kumar (1998) showed that if $(n\pi)r_n^2 = \ln n + c(n)$, then $P_{\Omega,1}(n, r_n) \to 1$ iff $c(n) \to +\infty$ as n goes to infinity. The result by Penrose (1998) implies a stronger result: If $(n\pi)r_n^2 = \ln n + \alpha$, then $P_1(n, r_n) = e^{-e^{-\alpha}}$ as n goes to infinity.

Topology Control

It is common to separate the network design problem from the management and control of the network in the communication network literature. The separation is very convenient and helps to significantly simplify these two tasks, which are already very complex on their own. Nevertheless, there is a price to be paid for this modularity because the decisions made at the network design phase may strongly affect the network management and control phase. In particular, if the issue of designing efficient routing schemes is not taken into account by the network designers, then the constructed network might not be suited for supporting a good routing scheme. For example, a backbone-like network topology is more suitable for a hierarchical routing method than a flat network topology.

Topology control is to select either a subset of wireless devices or a subset of communication links that will be used for the network operations such as routing. Notice that power assignment by use of a smaller transmission power of some nodes also will remove some links from the original communication graph induced by use of the maximum transmission power. Topology control often has more choices by intentionally not using some physical links for routing.

A wireless ad hoc network needs some special treatment because it intrinsically has its own special characteristics and some unavoidable limitations compared with those of wired networks. For example, wireless nodes are often powered by batteries only, and they often have limited memories. A transmission by a wireless device is often received by many nodes within its vicinity, which possibly causes signal interferences at these neighboring nodes. On the other hand, we can also utilize this property to save the communications needed to send some information. Unlike most traditional static communication devices, the wireless devices of MANETs often are moving during the communication. Therefore, it is more challenging to design a network topology for wireless ad hoc networks that is suitable for designing an efficient routing scheme to save

energy and storage memory consumption than it is to design one for the traditional wired networks. To simplify the question so we can derive some meaningful understanding of wireless ad hoc networks, we assume that the wireless nodes are quasi-static for a period of time. Then, in technical terms, the question we deal with is whether it is possible (and, if possible, then how) to design a network that is a subgraph of the original communication graph (mostly a UDG), such that it ensures both attractive network features such as bounded node degree, low-stretch factor, a linear number of links, and attractive routing schemes such as localized routing with guaranteed performances.

Unlike the wired networks that typically have fixed network topologies, each node in a wireless network can potentially change the network topology by adjusting its transmission range and/or selecting specific nodes to forward its messages, thus controlling its set of neighbors. The primary goal of topology control in wireless networks is to maintain network connectivity, optimize network lifetime and throughput, and make it possible to design power-efficient routing. Not every connected subgraph of the UDG plays the same important role in network designing. One of the perceptible requirements of topology control is to construct a subgraph such that the shortest path connecting any two nodes in the subgraph is not much longer than the shortest path connecting them in the original UDG. This aspect of path quality is captured by the *stretch factor* of the subgraph. A subgraph with a constant stretch factor is often called a *spanner*, and a spanner is called a *sparse spanner* if it has only a linear number of links. Here, we review and study how to construct a sparse network topology efficiently for a set of static wireless nodes.

Restricting the size of the network has been found to be extremely important in reducing the amount of routing information. The notion of establishing a subset of nodes that perform the routing has been proposed in many routing algorithms (Das and Bharghavan, 1997; Sinha *et al.*, 1999; Stojmenovic *et al.*, 2002; Wu and Li, 2001). These methods often construct a virtual backbone by using the connected dominating set (CDS) (Alzoubi *et al.*, 2002a; 2002c; Wan *et al.*, 2002a), which is often constructed from a dominating set or a maximal independent set. For a full review of the state of the art in constructing the backbone, see X.-Y. Li (2002).

The other imperative requirement for network topology control in wireless ad hoc networks is the fault tolerance. To guarantee a good fault tolerance, the underlying network structure must be k-connected for some $k > 1$; i.e., given any pair of wireless nodes, there need to be at least k disjoint paths to connect them. By setting the transmission range sufficiently large, the induced UDG will be k-connected without doubt. Because energy conservation is important to increase the life of the wireless device, then the question is how to find the minimum transmission range such that the induced UDG is multiply connected.

Limitations

For simplicity, it is traditionally assumed that the transmission region of each wireless node is a disk with unit radius. Here, a disk centered at a node u with radius r_u, denoted by $D(u, r_u)$, is the set of points whose distance to u is at most r_u. Thus, all nodes together define a UDG as a communication graph. However, graphs representing communication

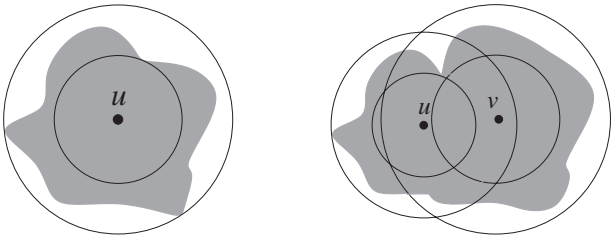


Figure 2.1 The transmission region of a node is modeled by a quasi-disk.

links are rarely specified as the UDG. Different nodes may have different transmission radii and, more important, the transmission region of a node is never a perfect disk. In other words, the main weakness of this point graph model (see Sen and Huson, 1997, for more details) is the assumption of perfectly regular radio coverage. Although this model is quite realistic in open-air and flat environments, ad hoc and sensor networks are likely to be used in various different situations, such as indoor or urban scenarios or under harsh conditions. In other words, in real-life situations, it is more likely that the radio coverage region is highly irregular because of the reflection influence of walls, buildings, and the interference with other infrastructures.

Considering this imperfect transmission region, previous routing algorithms, which guarantee packet delivery by use of some planar subgraph as network topology, are likely to fail because there might be no planar subgraph of the communication graph or some links might be missing. In the worst case, the communication graph could be very complicated. However, to have some meaningful study, including all these details in the network model, would make it extremely difficult and complex to study the performances of designed protocols and algorithms. For this reason, despite its limitations, the point graph model is still widely used in the study of wireless ad hoc network properties.

In addition to this point graph model, several enhanced models have been proposed in the literature to improve this. One model is the so-called quasi-disk-graph model (Barriere *et al.*, 2001; Moaveninejad *et al.*, 2005). Assume that each node u has a maximum transmission radius R_u and a minimum transmission radius r_u. These two thresholds depend on both the environment and the mobile hosts' technology. Thus, the transmission region of a node u is contained inside disk $D(u, R_u)$ and contains the disk $D(u, r_u)$. See Figure 2.1 for an illustration. If the Euclidean distance between two mobile hosts u and v exceeds the value $\min(R_u, R_v)$, they cannot communicate with each other directly (i.e., exchange messages). Conversely, two mobile hosts are always mutually reachable if their Euclidean distance is below the value $\min(r_u, r_v)$. Otherwise, they may or may not be mutually reachable.

2.4 The Wireless Interference Graph

Each terminal v_i also has an interference range $R_I(v_i)$ such that terminal v_j is interfered by the signal from v_i whenever $\|v_i - v_j\| \leq R_I(v_i)$ and v_j is not the intended receiver. Here, we assume that no node that transmits on a certain frequency can, at the same

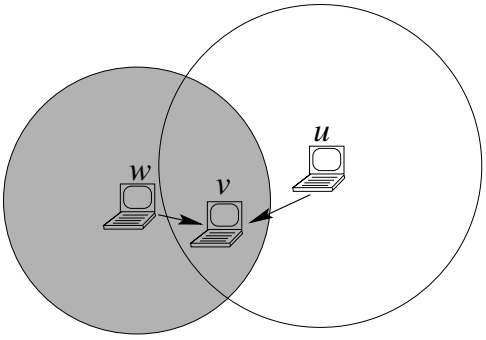


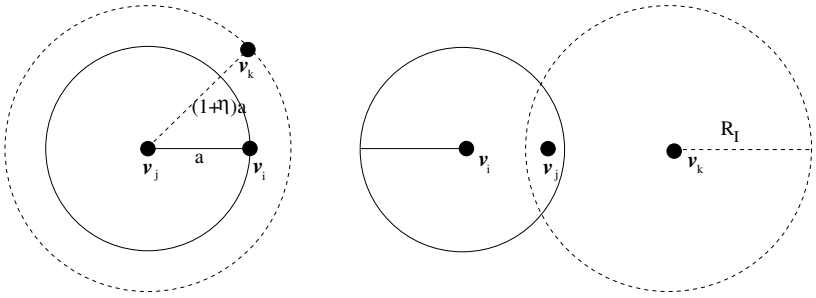
Figure 2.2 Interference happens at node v when the transmission region (denoted by the shaded disk) of node w intersects with the interference region of node u.

time, receive on the same frequency. Thus, we assume an interference occurs if the transmission region of one node (node w here) intersects with the interference region of another node (node u here). See Figure 2.2 for an illustration. Most researchers, for simplicity, treat the transmission region of a node as its interference region. However, this simplification is not accurate in practice. The interference range $R_I(v_i)$ is not necessarily the same as the transmission range $R_T(v_i)$. Typically, $R_T(v_i) < R_I(v_i) \leq cR_T(v_i)$ for some constant $c > 1$. We call the ratio between them the *interference–transmission ratio* for node v_i, denoted as $\gamma_i = [R_I(v_i)/R_T(v_i)]$. In practice, $2 \leq \gamma_i \leq 4$. For all wireless nodes, let $\gamma = \max_{v_i \in V}[R_I(v_i)/R_T(v_i)]$. Further, for a number of protocols, the actual simultaneous transmitting nodes must be separated by a distance called the *carrier-sensing range*. The carrier-sensing range for a node u is the largest range D such that a node that is of distance D away from u can still sense that the channel is busy when u is transmitting. Typically, the carrier-sensing range is larger than the interference range. For some theoretical studies, we need to use this carrier-sensing range to model the "interference" if two simultaneously transmitting nodes must be separated by at least their carrier-sensing range.

Two different types of interference have been studied in the literature: namely, *primary interference* and *secondary interference*. Primary interference occurs when a node transmits and receives packets at the same time; secondary interference occurs when a node receives two or more separate transmissions. Here, all transmissions could be intended for this node, or only one transmission is intended for this node (thus, all other transmissions are interference to this node). In addition to these interferences, there could have been some other constraints on the scheduling; e.g., the radio networks that deploy the IEEE 802.11 protocol with the request-to-send and clear-to-send (RTS/CTS) mechanism will pose some additional constraints. Several different interference models have been used to model the interferences in wireless networks. These are briefly reviewed in the following subsections.

Protocol-Interference Model (PrIM)

This model was first proposed in Gupta and Kumar (1999). In this model, a transmission by a node v_i is successfully received by a node v_j iff the intended destination



(a) Protocol Interference (b) Fixed-Power Protocol Interference

Figure 2.3 (a) Protocol-interference model: Node v_j will be interfered by node v_k when $\|v_k - v_j\| \leq (1+\eta)\|v_i - v_j\|$, where v_i is sending data to v_j and v_k is sending to other nodes. (b) Fixed-power-interference model: Node v_j will be interfered by node v_k when $\|v_k - v_j\| \leq R_I(v_k)$. Here, the dotted circle denotes the largest interference range of a node.

v_j is sufficiently apart from the source of any other simultaneous transmission; i.e., $\|v_k - v_j\| \geq (1+\eta)\|v_i - v_j\|$ for any node $v_k \neq v_i$. Here, the constant $\eta > 0$ models situations in which a guard zone is specified by the protocol to prevent a neighboring node from transmitting on the same channel at the same time. See Figure 2.3(a) for an illustration. This model *implicitly* assumed that each node v_k will adopt the power-control mechanism when it transmits signals. Simulation analysis (Gronkvist and Hansson, 2001) as well as analytical results (Behzad and Rubin, 2003) indicate that the protocol-interference model does not necessarily provide a comprehensive view of reality because of the aggregate effect of interference in wireless networks. However, it does provide some good estimations of interference and, most important, it enables a theoretical performance analysis of a number of protocols designed in the literature. This model was used in Kumar *et al.* (2005) to study the throughput optimization for wireless networks.

Fixed Power-Protocol-Interference Model (fPrIM)

We assume that a node will *not* dynamically change its power based on the intended receiver in a packet level. However, we do assume that each node v_i has its own fixed transmission power and thus a fixed transmission range $R_T(v_i)$. We also assume that each node v_k has an *interference range* $R_I(v_k)$ such that any node v_j will be interfered by the signal from v_k if $\|v_k - v_j\| \leq R_I(v_k)$ and node v_k is sending a signal to some node other than v_j. In other words, the transmission from v_i to v_j is viewed successful if $\|v_k - v_j\| > R_I(v_k)$ for every node v_k transmitting in the same time slot using the same channel. See Figure 2.3(b) for an illustration.

RTS/CTS Model

This model was also studied previously [e.g., Alicherry *et al.* (2005)]. When using the RTS/CTS mechanism, a transmitter first sends a RTS frame before sending a data frame. The intended receiver then responds with a CTS frame indicating that the transmitter can send the data frame. Within the CTS frame, the receiver provides a value in the duration